

Search for optimal measure for discriminating spike trains with different randomness

Keiji Miura*

*Department of Physics, Graduate School of Sciences,
Kyoto University Kyoto 606-8502, Japan and
Intelligent Cooperation and Control”,
PRESTO, JST,
c/o The University of Tokyo,
Chiba 277-8561, Japan*

Masato Okada†

*Department of Complexity Science and Engineering,
Graduate School of Frontier Sciences,
The University of Tokyo, Chiba 277-8561, Japan
Intelligent Cooperation and Control”, PRESTO, JST,
c/o The University of Tokyo, Chiba 277-8561, Japan
Laboratory for Mathematical Neuroscience,
RIKEN Brain Science Institute, Saitama 351-0198, Japan*

Shigeru Shinomoto‡

*Department of Physics, Graduate School of Sciences,
Kyoto University Kyoto 606-8502, Japan*

(Dated: February 9, 2008)

Abstract

We wish to discriminate spike sequences based on the degree of irregularity. For this purpose, we search for a rational expressions of quadratic functions of consecutive interspike intervals that efficiently measures spiking irregularity. Under natural assumptions, the functional form of the coefficient can be parameterized by a single parameter. The parameter is determined so as to maximize the mutual information between the distributions of coefficients computed for spike sequences derived from different renewal point processes. We find that the local variation of interspike intervals, L_V (Neural Comput. Vol. 15, pp. 2823-42, 2003), is nearly optimal for whose intrinsic irregularity is close to that of experimental data.

PACS numbers: Valid PACS appear here

*Electronic address: miura@ton.scphys.kyoto-u.ac.jp

†Electronic address: okada@k.u-tokyo.ac.jp

‡Electronic address: shinomoto@scphys.kyoto-u.ac.jp

I. INTRODUCTION

It is important to extract as much information as possible from spike sequences when looking for correlations between animal behaviors and neuronal activities [1, 2, 3, 4] or controlling prosthetic apparatuses by neuronal activities [5]. In many cases, however, only the mean firing rate is considered and the timing information is not taken into account. Consideration of detailed temporal structure of each spike train would help to decode brain signals more efficiently. We would like to propose a measure, which augments the information provided by the mean firing rate.

Coefficients that are functions of the interspike intervals (ISIs) are effective in detecting a spiking irregularity from a short spike train. For instance, the coefficient of variation, C_V , is widely adopted as a measure of the variance of ISIs [6, 7, 8, 9]. Recently, a measure of the local variation of interspike intervals, L_V , was proposed [10], as a natural extension of C_V which was designed to detect a stepwise variation of consecutive ISIs [11]. An analysis using L_V revealed that *in vivo* spike sequences are not uniformly random, but possess specific characteristics that vary among individual neurons. In addition, it was found that the neocortex consists of heterogeneous neurons that differ not only from one cortical area to another, but also from one layer to another in their spiking patterns [12].

In the present study, we try to modify L_V in an attempt to find a better measure for discriminating spike sequences based on the degree of irregularity. Namely, we examine rational expressions of quadratic functions of consecutive interspike intervals for suitability as coefficients for measuring spiking irregularity. Under reasonable assumptions, the functional form of the coefficient is found to be parameterized by a single parameter. The parameter is determined so as to maximize the mutual information between the distributions of coefficients computed for finite size sample sequences derived from different renewal gamma processes. It is found that L_V is not optimal for nearly random Poisson spike trains but optimal for more regular spike trains.

In Sec. II, we explain how we generated spike sequences with the same firing rate but different intrinsic irregularity. We show that a gamma distribution suffices for that purpose and that two parameters in the gamma distribution can be chosen as orthogonal coordinates. In Sec. III we explain L_V and compare it with C_V . We show that attractiveness of L_V stems from its symmetries. In Sec. IV we extend L_V and show that, under reasonable assumptions,

the extension of L_V can be parameterized by a single parameter. In Sec. V we explain how we determined the optimal value of the parameter using the maximization principle of mutual information. In Sec. VI, we determine the optimal value numerically. In Sec. VII, we describe our theory, developed using a Gaussian approximation, for explaining the results. In Sec. VIII, we discuss two non-stationary cases.

II. GENERATING SPIKE TRAINS WITH DIFFERENT RANDOMNESS

In this section, we explain how to generate spike trains with the same firing rate but different randomness.

There are many ways to generate spike trains artificially. For example, we can generate spike trains by using a network of spiking neuron models. However, we do not need to describe precise spike timing here, and a simple mechanism is desirable. Therefore, we assume that the mechanism is a renewal process and that the inter spike interval (ISI) follows a gamma distribution [6], which is described as

$$p(T) = \frac{1}{\Gamma(\kappa)} \left(\frac{\kappa}{\mu} \right)^\kappa T^{\kappa-1} e^{-\frac{\kappa}{\mu}T}, \quad (1)$$

where T denotes an ISI. We generate ISIs from the distribution and align them to make a spike train. The mean and variance of the ISIs are

$$\begin{cases} Ex(T) = \mu \\ Var(T) = \frac{\mu^2}{\kappa}. \end{cases} \quad (2)$$

The mean firing rate is obtained by taking the inverse of the mean ISI [13]. The κ is a shape parameter; $\kappa = 1$ corresponds to an exponential distribution, and, as κ increases, the distribution approaches a normal distribution. The exponential distribution corresponds to a Poisson process in which the firing rate (hazard function) is constant with time independent of the previous firing time. The spike train looks random. As κ increases, the variance of the ISIs decreases, and the ISIs become regular.

Our goal is to find an optimal measure for discriminating two spike trains with different randomness independent of their mean firing rates. A gamma distribution is suitable for that purpose. First, we can control the mean firing rate and randomness independently by changing the two parameters (μ and κ) in the distribution. Next, experimental data can

be well fitted by the distribution. For example, Baker et al. showed that the spike patterns recorded from primary and supplementary motor areas are explicable using a gamma distribution [14].

We can transform the parameters in a gamma distribution arbitrarily. For example, we can transform the parameters into (α, λ) :

$$\begin{cases} \alpha = \kappa, \\ \lambda = \frac{\kappa}{\mu}. \end{cases} \quad (3)$$

The gamma distribution in this coordinate can be written as

$$p(T) = \frac{\lambda^\alpha}{\Gamma(\alpha)} T^{\alpha-1} e^{-\lambda T}, \quad (4)$$

where λ is a scale parameter. The mean and variance of the ISIs can be written as functions of α and λ as

$$\begin{cases} Ex(T) = \frac{\alpha}{\lambda}, \\ Var(T) = \frac{\alpha}{\lambda^2}. \end{cases} \quad (5)$$

Thus, there are many ways of writing (parameterizing) a gamma distribution. We used the expression shown as Eq. (1) because μ corresponds to the mean ISI and κ is orthogonal to it in the sense of information geometry [15, 16]. The proof is shown in APPENDIX A. We call the parameters of a gamma distribution coordinates because we regard the family of gamma distributions as manifold. We would like to define randomness as information orthogonal to the firing rate. Therefore, we regard κ as randomness in what follows. We generate spike trains having different intrinsic randomness by using the gamma distributions with different values of κ .

III. L_V AND C_V

The measure of local variation proposed by Shinomoto et al. [10] is defined as

$$L_V = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{3(T_i - T_{i+1})^2}{(T_i + T_{i+1})^2}, \quad (6)$$

where T_i denotes the i -th ISI in a spike train. The coefficient “3” is multiplied so that $\overline{L_V}$ is 1 for a Poisson process. L_V is large when consecutive ISIs differ. It is dimensionless and invariant if all the ISIs are multiplied by a constant. The conventional C_V is defined as [11]

$$C_V \equiv \frac{\sqrt{Var(T)}}{Ex(T)}. \quad (7)$$

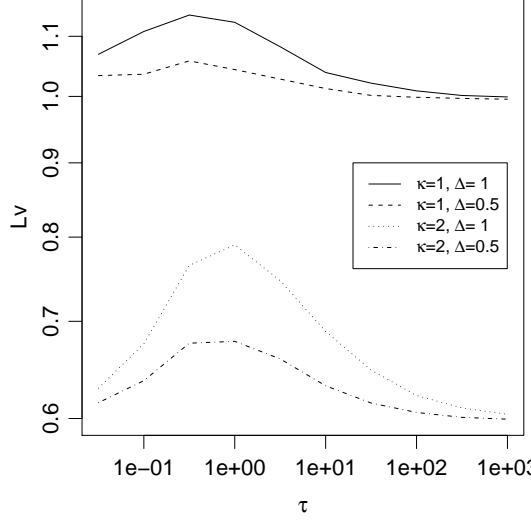


FIG. 1: L_V for doubly stochastic gamma process with various values of time constant τ and rate amplitude Δ .

Next we examine the difference between L_V and C_V and calculate L_V and C_V for the rate modulated gamma process.

We define a rate modulated gamma process as an extension of a gamma distribution where the firing rate, $\lambda(t)(= \frac{1}{\mu(t)})$, is time-dependent while κ is time-independent. The spikes for the rate modulated gamma process are generated as follows [7, 17]. Note that we consider only the case of integer κ . A spike is generated with probability $\lambda(t)dt$ for every small time step, dt . To be precise, we generate a uniform random number and if it is less than $\lambda(t)dt$, we generate a spike at that time step. For the case where κ is larger than 1, we keep every κ -th spike and remove the others. What is left is the desired sequence. In fact, for the case where λ is constant over time, the spike sequence generated in this way is equivalent to that generated from a renewal gamma distribution with $\mu = \frac{1}{\lambda}$.

Here we consider a doubly stochastic gamma process whose firing rate obeys the Ornstein-Uhlenbeck process [18]. We assume the firing rate, λ , satisfies

$$\frac{d\lambda}{dt} = -\frac{\lambda - \lambda_0}{\tau} + \Delta\sqrt{\frac{2}{\tau}}\xi(t), \quad (8)$$

where ξ is Gaussian white noise, $\langle \xi(t) \rangle = 0$, and $\langle \xi(t), \xi(t') \rangle = \delta(t - t')$.

Fig. 1 and Fig. 2 show L_V and C_V with $\lambda_0 = 1$ for various values of time constant τ and rate amplitude Δ . For simplicity, we consider sufficiently long spike sequences and assume that the values of L_V and C_V converge. Fig. 1 shows that in the limit of a large time constant,

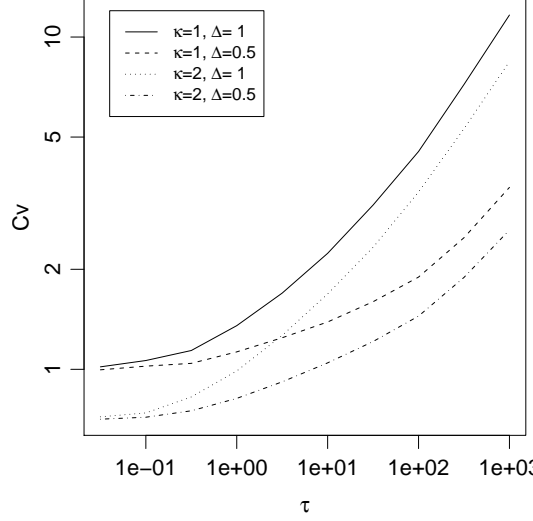


FIG. 2: C_V for doubly stochastic gamma process with various values of time constant τ and rate amplitude Δ .

the values of L_V converge to the value for the stationary case. This means that the value of L_V does not depend on the amplitude of the firing rate and has one-to-one correspondence with κ in this limit. Fig. 2 shows that C_V depends on both κ and Δ and does not have one-to-one correspondence with κ . Therefore, L_V is better than C_V for discriminating the intrinsic randomness of spike sequences.

This attractive property seems to stem from the fact that L_V is the sum of the dimensionless terms of consecutive interspike intervals. By “dimensionless” we mean that the numerator and denominator have the same dimension. Every term in L_V is normalized locally by the average of two consecutive interspike intervals instead of the global average. Intuitively, because the firing rates for two consecutive interspike intervals can be regarded as the same in the slow limit, terms should be the same as those for the stationary case. On the other hand, C_V is the variance around the global mean of the ISIs and can be large for both the case where the firing rate fluctuates significantly and the case where the intrinsic randomness is large. Therefore, we cannot distinguish the two cases based on the value of C_V .

IV. MEASURE OF LOCAL VARIATION

We extend L_V without losing its attractive property described in the previous section and find a better measure of intrinsic randomness. We do this by focusing on the ISI statistics and imposing three symmetry conditions: (1) time translation invariance, (2) time-scale transformation invariance, and (3) time inversion invariance.

We assume the randomness of a spike train is constant over time and define the extended L_V as

$$\widetilde{L}_V = \frac{1}{n-1} \sum_{i=1}^{n-1} f(T_i, T_{i+1}), \quad (9)$$

where T_1, T_2, \dots, T_n are the observed ISIs and $f(x, y)$ does not depend on i explicitly. This form guarantees invariance under time translation ($i \rightarrow i+1$) if n is infinite. Next, we assume that f is invariant under the time-scale transformation ($T \rightarrow kT$). This requires that the denominator and numerator of f have the same dimension. For simplicity, we assume that the dimension is two, so f can be written as

$$f(x, y) = \frac{c_1 x^2 + c_2 xy + c_3 y^2}{c_4 x^2 + c_5 xy + c_6 y^2}, \quad (10)$$

which includes the original L_V as a specific case. In addition, because we do not distinguish increases from decreases in the firing rate in terms of randomness, we impose time inversion invariance and require

$$f(x, y) = f(y, x). \quad (11)$$

Thus, f can be written as

$$f(x, y) = \frac{c_1 x^2 + c_2 xy + c_1 y^2}{c_4 x^2 + c_5 xy + c_4 y^2}. \quad (12)$$

Note that the absolute value of L_V does not matter in discriminant analysis, and we can add (or multiply by) a constant to f . Then, without loss of generality, f can be written as

$$f(x, y) = \frac{xy}{x^2 + c_5 xy + y^2}. \quad (13)$$

In addition, we can rewrite the denominator using $c = c_5 + 2$:

$$f(x, y) = \frac{xy}{(x - y)^2 + cxy}. \quad (14)$$

Because each term in the denominator is non-negative, the necessary and sufficient condition that the denominator always be positive is $c > 0$.

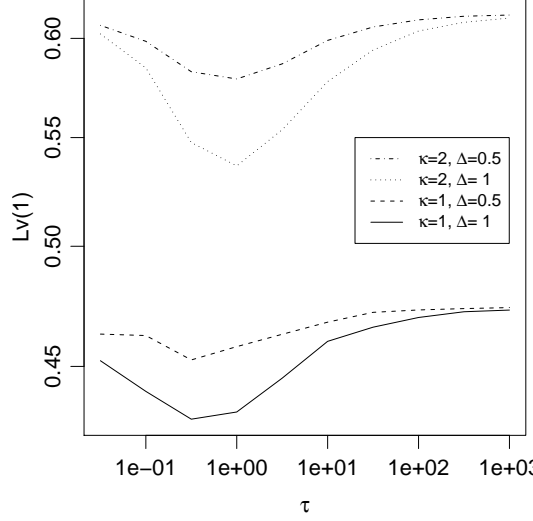


FIG. 3: $\widetilde{L}_V(1)$ for doubly stochastic gamma process with various values of time constant τ and rate amplitude Δ .

As a result, \widetilde{L}_V can be written as

$$\widetilde{L}_V(c) = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{T_i T_{i+1}}{(T_i - T_{i+1})^2 + c T_i T_{i+1}}. \quad (15)$$

Note that the original L_V corresponds to the case of $c = 4$. In this way, the measures satisfying the symmetries have only one degree of freedom and can be parametrized by a single parameter.

The \widetilde{L}_V should have one-to-one correspondence to κ like L_V because of its symmetries. In fact, it has the same values as those for the stationary case in the limit of a large time constant for the doubly stochastic gamma process. Fig. 3 shows that $\widetilde{L}_V(1)$ is independent of the rate amplitude, Δ , and is a function of κ in the limit. The results for other values of c , for instance $\widetilde{L}_V(16)$, remain the same.

Thus, $\widetilde{L}_V(c)$ has one-to-one correspondence with κ . However, this is not sufficient to make it a good measure. We previously have considered only spike sequences with infinite length. However, in practical experimental situations, data sizes are limited, and $\widetilde{L}_V(c)$ varies widely by trial around the mean. Similarly, if spike sequences are generated using a gamma distribution, $\widetilde{L}_V(c)$ varies by trial for the finite spike sequence. In the discrimination of intrinsic randomness, roughly speaking, the smaller the variance, the higher the hitrate. Thus, we next search for an optimal value of parameter c , where the variance is the smallest.

V. MUTUAL INFORMATION MAXIMIZATION PRINCIPLE

We use the mutual information maximization principle to determine an optimal measure. We assume that the firing rate is constant over time and spike sequences are generated by a gamma distribution, as shown in Sec. II. As shown in Sec. III, \widetilde{L}_V does not depend on μ . Here we set $\mu = 1$. We consider the stationary case because it is tractable and can be regarded as the slow change limit of the firing rate. We show in Sec. VIII that the optimal value of c for the nonstationary case does not differ significantly from that for the stationary case.

The optimal parameter value is determined so as to maximize the mutual information between the coefficients and randomness. Here we assume that a spike train consists of 100 ISIs because this is the typical length available from laboratory experiments. \widetilde{L}_V can be computed for a spike train, and the value of \widetilde{L}_V varies among spike trains. Even if spike trains are generated from the same distribution, the values of \widetilde{L}_V can differ because the length of a spike train is finite. As a result, the distribution of \widetilde{L}_V can be obtained for one parameter set of the gamma distribution. Thus, two distributions can be obtained from two types of spike trains. The mutual information can be computed from the two distributions. The bigger the mutual information, the better randomness (κ) can be discriminated based on the observed \widetilde{L}_V .

Mutual information is calculated as follows. Spike trains are generated from two gamma distributions with equal probability, $\frac{1}{2}$. The two distributions have different κ . All the ISIs in a spike train are generated by using the same distribution. We denote the distribution of \widetilde{L}_V generated from the $i(= 1, 2)$ -th gamma distribution as $p(x|i)$; $p(x)(= \frac{1}{2}p(x|1) + \frac{1}{2}p(x|2))$ represents the distribution of \widetilde{L}_V with no distinction of the source. The entropy is defined as

$$H = - \int p(x) \ln p(x) dx. \quad (16)$$

The noise entropy is defined as

$$H_n = -\frac{1}{2} \int p(x|1) \ln p(x|1) dx - \frac{1}{2} \int p(x|2) \ln p(x|2) dx. \quad (17)$$

The mutual information is the difference,

$$I_m = H - H_n. \quad (18)$$

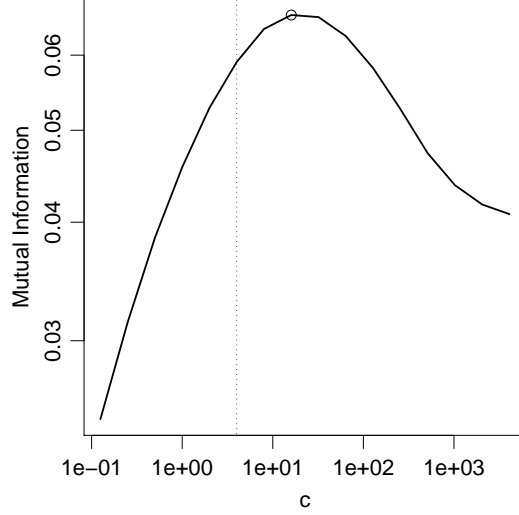


FIG. 4: Mutual information with $\kappa_1 = 1, \kappa_2 = 1.1$. Open circle denotes peak. Dotted line is for $c = 4$ corresponding to original L_V . Mutual information has a peak with c larger than 4.

The mutual information is the reduction in uncertainty about the spike trains due to the knowledge of \widetilde{L}_V . Mutual information is 0 if two distributions of \widetilde{L}_V are identical so that they cannot be distinguished. Mutual information is 1 if two distributions of \widetilde{L}_V have no overlap, and only one sample of \widetilde{L}_V is needed to distinguish them.

In the next section, we will show the results of a Monte Carlo simulation. We calculated mutual information as a function of c for various sets of randomness, κ_1 and κ_2 .

VI. RESULTS

Fig. 4 shows the mutual information with $\kappa_1 = 1$ and $\kappa_2 = 1.1$; κ_1 and κ_2 are the shape parameters of two gamma distributions and c is the parameter in $\widetilde{L}_V(c)$. We set the number of ISIs per spike train, n , to 100. The mutual information has a peak, whose location we denote by c_{peak} . The vertical line represents $c = 4$, which corresponds to the original L_V . Since $c_{peak}(\approx 16)$ is bigger, the optimal coefficient in this case is not the original L_V but $\widetilde{L}_V(16)$. However, c_{peak} depends on various parameters, and we will examine how it depends on the number of ISIs per spike train, κ_1 and κ_2 , in what follows. We can use the maximum likelihood estimator of κ as a measure instead of L_V , and the peak value of the mutual information for κ is 0.097. (For the maximum likelihood estimator, see Appendix B.) The peak value for L_V is about 0.066, which is smaller than that for the maximum likelihood

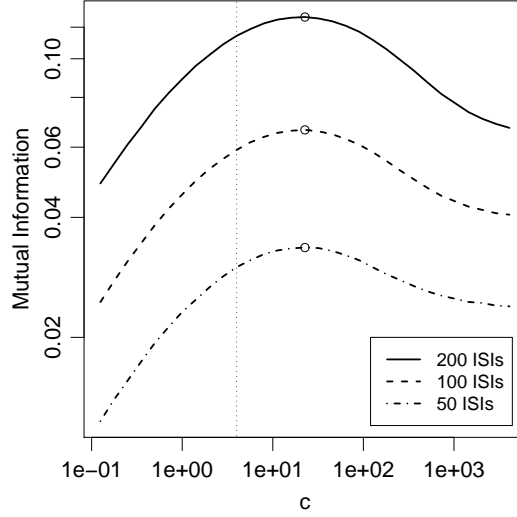


FIG. 5: Mutual information for various numbers of ISIs per spike sequence with $\kappa_1 = 1, \kappa_2 = 1.1$. Open circles denote peaks. Dotted line is for $c = 4$ corresponding to original L_V . Peak location almost does not depend on number of ISIs.

estimator. We nonetheless use L_V because the maximum likelihood estimator cannot be applied to the nonstationary case. In the cases where the firing rate is time-dependent, the mutual information for L_V can be much higher than that for the maximum likelihood estimator, as we will show in Sec. VIII.

Fig. 5 shows the mutual information for various numbers of ISIs per spike train. While the mutual information increases with the number of ISIs, the peak location remains almost the same. Although we show only the case for $\kappa_1 = 1, \kappa_2 = 1.1$, the other cases have similar results. Therefore, we set the number of ISIs per spike train to 100.

Fig. 6 shows the mutual information with $\kappa_1 = 1$ and various κ_2 . As $d\kappa (= \kappa_2 - \kappa_1)$ increases, the mutual information approaches 1. The peak location remains almost unchanged ($c_{peak} \approx 16$). For $\kappa_2 = 3.2$, the mutual information is almost 1, and the two distributions are completely distinguishable. In general, c_{peak} largely depends on κ_1 and is almost independent of κ_2 .

Fig. 7 shows the mutual information with $\kappa_2 = 1.3\kappa_1$ and various κ_1 . The peak location decreases with increasing κ_1 . For $\kappa_1 = 16$, the original L_V is nearly optimal ($c_{peak} \approx 4\sqrt{2}$). Since reported experimental data can be well fitted by a gamma distribution with $\kappa \approx 16$ [14], L_V seems to be optimal not for the Poisson data but for the experimental data.

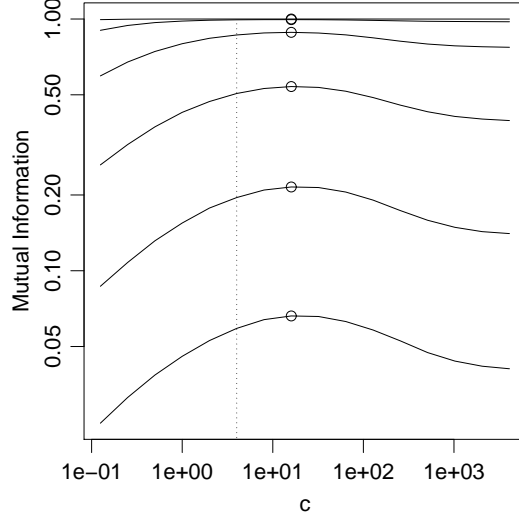


FIG. 6: Mutual information for various κ_2 with $\kappa_1 = 1$. Lines are for $\kappa_2 = 0.1, 0.2, 0.4, 0.8, 1.6$ and 3.2 from below. Open circles denote peaks. Dotted line is for $c = 4$ corresponding to original L_V . Peak location almost does not depend on κ_2 .

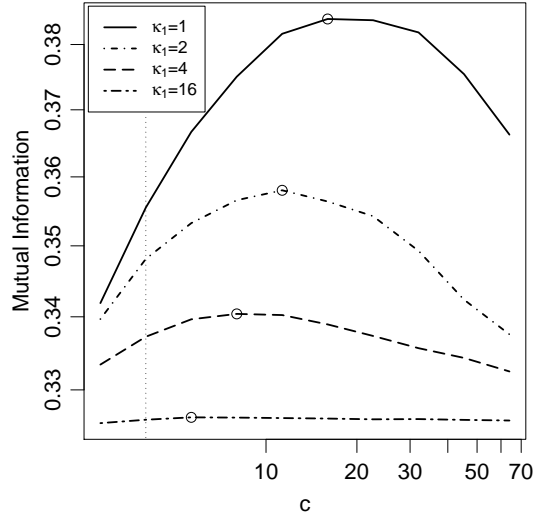


FIG. 7: Mutual information for various κ_1 with $\kappa_2 = 1.3\kappa_1$. Open circles denote peaks. Dotted line is for $c = 4$ corresponding to original L_V . Peak location decreases as κ_1 increases.

VII. THEORETICAL ANALYSIS

In this section we analyze the property of the mutual information theoretically. For simplicity, we do two approximations.

First, we consider the limit of a large number of ISIs per spike train and approximate

the distribution of L_V by using the normal distribution. Although this approximation is not good for $c \approx 0$, the peak location is far larger than 0 and can be discussed within this approximation.

In addition, we consider the limit of small $d\kappa$. In the limit, the mutual information can be written using the Fisher information [19] as

$$I_m = \frac{1}{8} J(p(x, \kappa)) d\kappa^2, \quad (19)$$

where the Fisher information is defined as

$$J(p(x, \kappa)) = \text{Ex}\left(\left(\frac{d \log p(x, \kappa)}{d\kappa}\right)^2\right). \quad (20)$$

This relation can be easily derived. We represent two L_V distributions as

$$p_1(x) = \frac{1}{\sqrt{2\pi\sigma(\kappa)^2}} e^{-(x-m(\kappa))^2/2\sigma(\kappa)^2} \quad (21)$$

and

$$p_2(x) = \frac{1}{\sqrt{2\pi\sigma(\kappa+d\kappa)^2}} e^{-(x-m(\kappa+d\kappa))^2/2\sigma(\kappa+d\kappa)^2}. \quad (22)$$

Inserting these equations into the definition of the mutual information and expanding by $d\kappa$ to the second order lead to the relation.

The Fisher information can be explicitly written as

$$J = \frac{m'(\kappa)^2 + 2\sigma'(\kappa)^2}{\sigma(\kappa)^2}. \quad (23)$$

Because σ^2 is inversely proportional to N , σ can be written as

$$\sigma = \frac{\sigma_0}{\sqrt{N}}. \quad (24)$$

The Fisher information can then be approximated as

$$\begin{aligned} J/N &= \frac{m'(\kappa)^2 + 2\frac{1}{N}\sigma_0'(\kappa)^2}{\frac{1}{N}\sigma_0(\kappa)^2} \frac{1}{N} \\ &\simeq \frac{m'(\kappa)^2}{\sigma_0(\kappa)^2}, \end{aligned} \quad (25)$$

where m' and σ_0 depend on only κ and c . As a result, the mutual information can be written as

$$I_m = \frac{1}{8} \frac{m'(\kappa, c)^2}{\sigma_0(\kappa, c)^2} N d\kappa^2. \quad (26)$$

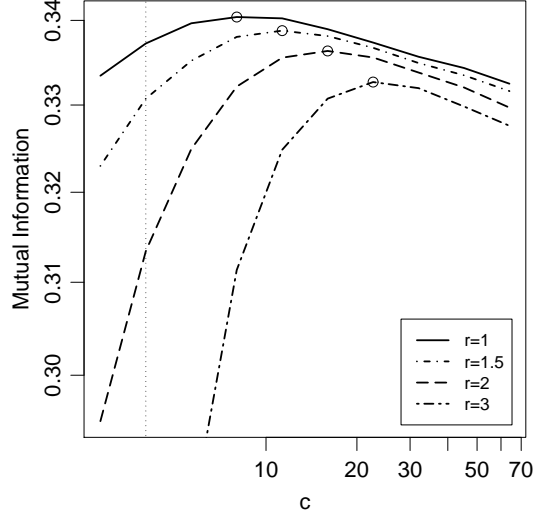


FIG. 8: Mutual information for monotonically decreasing firing rate for various r with $\kappa_1 = 4$ and $\kappa_2 = 5.2$.

Thus, I_m is proportional to N and $d\kappa^2$. The c dependency of I_m stems from only m' and σ_0 . Therefore, when N or $d\kappa$ changes, the absolute value of the mutual information changes while the peak location does not change. This is consistent with our numerical results in which the peak location did not depend on N and $d\kappa$. The peak location can be explained by an interplay of m' and σ_0 . However, m and σ_0 cannot be predicted solely by our theory. Numerical calculations are necessary for finding the peak location.

VIII. NONSTATIONARY CASE

We considered the discrimination of randomness for the stationary gamma process in the previous sections. However, it has been reported that experimental data can be explicable by the rate-modulated gamma process [14]. Therefore, we consider the rate-modulated gamma process in this section. We show two simple cases in which the firing rate decreases monotonically or changes stepwise.

A. monotonically decreasing firing rate

We consider a simple rate-modulated case and show that the peak location of the mutual information, c_{peak} , tends not to change if the firing rate fluctuates significantly. We generate

the ISIs by again using a gamma distribution. We assume that the mean ISI increases monotonically. For simplicity, we set $\mu_i = r^i$, where μ_i denotes the mean of the i -th ISI. We simply align n ISIs to make a single spike train as before. The value of κ does not change within the train. The mutual information is calculated for two spike trains with different values of κ .

Fig. 8 shows the mutual information for $\kappa_1 = 4$ and $\kappa_2 = 5.2$. The peak location decreases gradually from the stationary value as r increases. However, only extreme cases, in which the firing rates decrease more than 1.5 times one after another, are plotted. For realistic cases, c_{peak} changes only slightly. For example, the ratio between the last and first mean ISI is

$$\frac{\mu_n}{\mu_1} = r^{n-1}, \quad (27)$$

and the ratio is 2.678033 for $r = 1.01$ and $n = 100$ and 12527.83 for $r = 1.1$ and $n = 100$. This illustrates that the 1.5 used for r is extremely large. Similar results were obtained for different values of κ , so c_{peak} apparently tends not to change even if the firing rate fluctuates. This result is not restricted to the decreasing firing rate case. For example, the mean \widetilde{L}_V remains the same if a small and a large mean ISI appear alternately instead of the firing rate increasing monotonically. It thus appears that the peak location of the mutual information is almost independent of the firing rate if the variation in the firing rate is small.

B. stepwise changing firing rate

So far we have considered only $\widetilde{L}_V(c)$. However, the maximum likelihood estimator, $\hat{\kappa}$, should be better for the stationary case. Here we consider the case of a stepwise changing firing rate to show why we favor $\widetilde{L}_V(c)$ nonetheless. In a word, $\hat{\kappa}$ is not good for the nonstationary case because it is the maximum likelihood estimator for the stationary case, as shown in Appendix B. In principle, the firing rate at every small time bin can be estimated for the nonstationary case. However, doing so requires many spike sequences and the firing rate profile must be the same for all the sequences. Therefore, it is not practical for many realistic cases. Instead we consider simple measures like $\widetilde{L}_V(c)$ and $\hat{\kappa}$ even in the nonstationary case. In this section, we compare L_V and $\hat{\kappa}$ for the nonstationary case.

Consider the case in which the firing rate is stepwise increasing, as shown in the Fig. 9. At time $t = 50$, it shifts from 1 to λ_2 . Two types of spike trains, with $\kappa_1 = 16$ and $\kappa_2 = 20$,

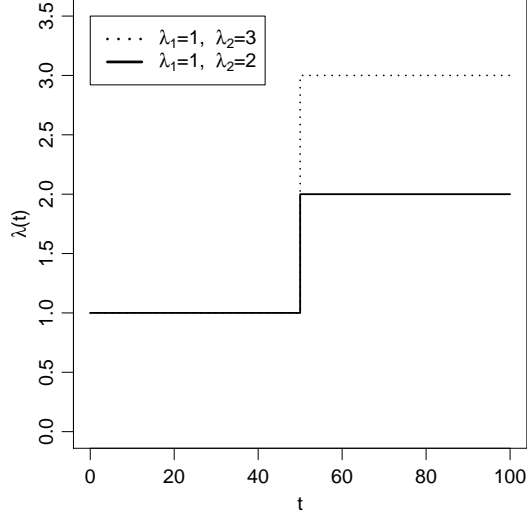


FIG. 9: Schematic diagram of stepwise increasing firing rate. Firing rate shifts from 1 to λ_2 at $t = 50$.

are generated based on the firing rate profile. Fig. 10 shows the mutual information for these trains when L_V or $\hat{\kappa}$ is used as a measure. The mutual information for L_V is independent of λ_2 in the limit of a large number of ISIs per train. The reason is that L_V is independent of the firing rate for the stationary case and in this case the firing rate is constant over time except for the discontinuous point. The contribution of the term in L_V that cross the discontinuous point is $O(1/n)$ and is small if the number of ISIs is large enough. We plotted the value for the stationary case, neglecting the contribution for simplicity. On the other hand, the mutual information for $\hat{\kappa}$ decreases as λ_2 increases. For example, when the firing rate increases 1.5 times, the mutual information for L_V is larger than that for $\hat{\kappa}$.

Thus, for a stepwise increasing firing rate, L_V is better than $\hat{\kappa}$. This type of sudden change can be observed when a visual stimulus is presented to a monkey at a given time. The result remains almost the same for the stepwise firing rate with multiple discontinuous points in the limit of a large number of ISIs. In addition, $\hat{\kappa}$ depends on both κ and the amplitude of the firing rate, as shown in Sec. II for C_V . Therefore, L_V is a better measure of intrinsic randomness.

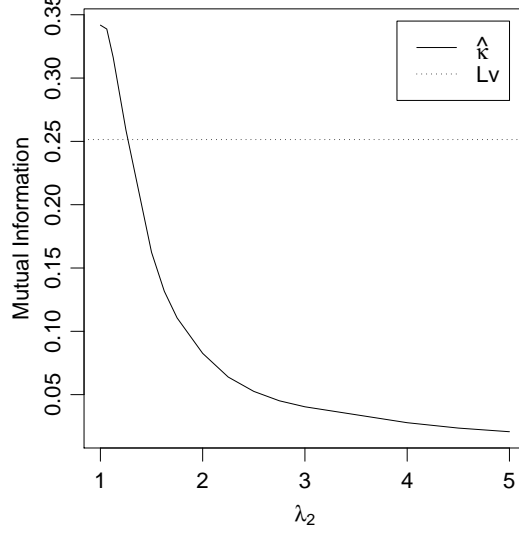


FIG. 10: Mutual information for stepwise increasing firing rate with $\kappa_1 = 16$ and $\kappa_2 = 20$.

IX. SUMMARY AND DISCUSSION

In this study, we sought a measure more effective than the local variation of interspike intervals, L_V , in discriminating spike trains based on the degree of intrinsic spiking irregularity.

We first compared characteristics of the conventional coefficient of variation, C_V , and the local variation, L_V . The coefficient of variation, C_V , measures a global variability of ISIs, and therefore depends on not only the local irregularity of ISIs but also the rate fluctuation, which would naturally manifest itself in *in vivo* neuronal spiking conditions. In contrast, the local variation, L_V , measures only a stepwise variability of ISIs, and therefore does not depend significantly on a rate fluctuation. It was revealed that L_V is superior to C_V in detecting some intrinsic spiking irregularity specific to individual neurons *in vivo* [10, 12].

For a spike train of a finite number of ISIs derived from a given point process, the value of L_V as well as C_V varies from trial to trial. The goodness of a coefficient is quantified by its narrow distribution of values among spike trains derived from the same point process and the small overlap of this distribution with the distribution obtained from spike trains derived from a different point process. In other words, we sought a new coefficient that maximizes the mutual information between spike sequences created from different renewal gamma processes.

For this purpose, we adopted a rational expression of quadratic functions of consecutive

interspike intervals that is the same form as L_V , and searched for the optimal parameter of the coefficient. The optimal parameter of the coefficient depends on the choice of the point processes that are to be discriminated. It was found that the original L_V is not optimal for near random (Poisson) point processes, but is optimal for more regular spike trains. In this way, if we have preliminary knowledge of the spiking irregularities of the point processes, we are able to propose a better coefficient than the original L_V for the purpose of discriminating spike trains.

We generated spike sequences entirely by using a stationary or rate-modulated gamma process. The reason is as follows. The Poisson process, in which the firing rate is represented as a function of time from stimulus onset, is widely used in spike data analysis [20]. However, the statistical properties of spike sequences cannot be fully captured by the rate-modulated Poisson process [21, 22, 23, 24]. In other words, spike probability depends on the past spike times due to the so-called refractory period. A gamma process is a Poisson process with an additional parameter representing a kind of refractory period. Baker et al. showed that the spike pattern recorded from primary and supplementary motor areas is explicable by a gamma process [14].

We considered only mutual information as a measure for discriminating two spike trains. However, the Kullback-Leibler divergence $D(p_1, p_2)$ is a well-known measure of the dissimilarity of two distributions, too. It is also proportional to the Fisher information, $D = \frac{1}{2}J(p(x, \kappa))d\kappa^2$, under the same approximation as described in Sec. VII. Note that the coefficient is $\frac{1}{2}$ instead of $\frac{1}{8}$, as seen in Eq. (19), for the mutual information. However the coefficient is irrelevant to the peak location. Thus, the Kullback-Leibler divergence leads to the same results as mutual information. Nonetheless, we used mutual information because it is symmetrical in terms of two distributions. The Kullback-Leibler divergence is not symmetrical. Its value changes if the two distributions are interchanged. The Kullback-Leibler divergence becomes symmetrical in the limit of a small difference of two distributions, where it is proportional to the Fisher information.

In previous studies, various measures were computed for mathematical models [13, 25, 26]. However, the focus was only on the expectations for the measures. In discrimination tasks, the variance of a measure is more important than the expectation. For example, consider the case in which the expectations of a measure for two different types of spike sequences differ considerably. If the variances are very large, discriminating the two sequences is difficult. In

addition, if the definition of a measure is changed, for example, multiplied or added to by a constant, the expectation changes, but the mutual information never changes. Therefore, in this paper we focused on the variance and searched for the measure that maximizes the mutual information.

APPENDIX A: ORTHOGONAL COORDINATES FOR GAMMA DISTRIBUTION

We show that κ and μ are orthogonal coordinates in the sense of information geometry. The theory of information geometry is described elsewhere [15, 16], and there are applications to neuroscience [27, 28, 29].

For the purpose of proving the orthogonality, it suffices to demonstrate that the Fisher information matrix is diagonal. The Fisher information matrix is defined as

$$g_{ij} = \int_0^\infty \frac{\partial \log p(T)}{\partial \xi^i} \frac{\partial \log p(T)}{\partial \xi^j} p(T) dT, \quad (\text{A1})$$

where $\xi^1 = \mu$ and $\xi^2 = \kappa$. The log-likelihood can be written as

$$\log p(T) = \kappa \log\left(\frac{\kappa}{\mu}\right) + (\kappa - 1) \log T - \log \Gamma(\kappa) - \frac{T\kappa}{\mu}. \quad (\text{A2})$$

The derivatives of the log-likelihood are

$$\frac{\partial \log p(T)}{\partial \mu} = -\frac{\kappa}{\mu} + \frac{T\kappa}{\mu^2} \quad (\text{A3})$$

and

$$\frac{\partial \log p(T)}{\partial \kappa} = \log \frac{\kappa}{\mu} + 1 + \log T - \psi(\kappa) - \frac{T}{\mu}, \quad (\text{A4})$$

where $\psi(\kappa) = (\log \Gamma(\kappa))'$. The matrix elements can be written as

$$g_{\mu\mu} = \frac{\kappa}{\mu^2}, \quad (\text{A5})$$

$$g_{\mu\kappa} = g_{\kappa\mu} = 0, \quad (\text{A6})$$

$$g_{\kappa\kappa} = \psi(\kappa)' - \frac{1}{\kappa}. \quad (\text{A7})$$

Thus, the Fisher information matrix is diagonal at every point. According to the theory of information geometry, it is always possible to choose orthogonal coordinates for an exponential family of distributions that includes the gamma distribution as a specific case.

The Fisher information matrix has the meanings described below. When μ and κ are estimated from a finite number of samples, the estimated values are not necessarily the same as the true value. The value of the maximum likelihood estimator varies depending on the sample sets, and its variation around the true value can be approximated by a normal distribution whose variance is the inverse of the Fisher matrix if the sample size is sufficiently large [19]. Thus, the diagonality of the Fisher matrix means that the variations in the maximum likelihood estimators of μ and κ are uncorrelated.

APPENDIX B: MAXIMUM LIKELIHOOD ESTIMATION FOR GAMMA DISTRIBUTION

Let T_1, T_2, \dots, T_n be observed ISIs. We would like to estimate the true values of μ and κ from them. The log-likelihood is defined as

$$l \equiv \ln(p(T_1)p(T_2)\dots p(T_n)) \quad (\text{B1})$$

and can be written as

$$l = n\kappa \ln \frac{\kappa}{\mu} - n\Gamma(\kappa) + (\kappa - 1) \sum \ln T_i - \frac{\kappa}{\mu} \sum T_i. \quad (\text{B2})$$

The maximum likelihood estimators must satisfy both $\frac{\partial l}{\partial \mu} = 0$ and $\frac{\partial l}{\partial \kappa} = 0$. The derivatives of the log-likelihood are

$$\frac{\partial l}{\partial \mu} = \frac{\kappa}{\mu^2} \sum T_i - n \frac{\kappa}{\mu} \quad (\text{B3})$$

and

$$\frac{\partial l}{\partial \kappa} = \sum \ln T_i - \frac{1}{\mu} \sum T_i + n \ln \frac{\kappa}{\mu} + n - n\psi(\kappa). \quad (\text{B4})$$

Then, $\hat{\mu}$ can be explicitly obtained as

$$\hat{\mu} = \frac{1}{n} \sum T_i, \quad (\text{B5})$$

and $\hat{\kappa}$ must satisfy

$$\frac{1}{n} \sum \ln T_i - \ln \frac{1}{n} \sum T_i = \psi(\hat{\kappa}) - \ln \hat{\kappa}, \quad (\text{B6})$$

where $\psi(\hat{\kappa}) = (\log \Gamma(\hat{\kappa}))'$. This equation cannot be solved explicitly for $\hat{\kappa}$. However, the right side of the equation is a monotonic function of $\hat{\kappa}$, and we can obtain $\hat{\kappa}$ by numerical iteration.

Instead of a lengthy numerical iteration, we can use the moment estimator. According to Eq. (2), we can estimate the true κ from the sample mean and variance:

$$\kappa = \frac{Ex(T)^2}{Var(T)}. \quad (\text{B7})$$

In fact, the right side of Eq. (B7) can be rewritten as $\frac{1}{C_V^2}$. Thus, we can regard C_V as a moment estimator. However, the moment estimator is worse than the maximum likelihood estimator, especially when κ is close to 1 [6]. Nevertheless, it is good as a first approximation, and we can use it as the initial value of the numerical iteration in maximum likelihood estimation.

Another way to avoid numerical iteration is to use the left side of Eq. (B6) as a measure. In discriminant analysis, we do not need to estimate κ because the left side of Eq. (B6) has one-to-one correspondence with $\hat{\kappa}$ and has the same information as $\hat{\kappa}$.

-
- [1] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, *Science* **233**, 1416 (1986).
 - [2] Y. Miyashita and H. S. Chang, *Nature* **331**, 68 (1988).
 - [3] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic, *J. Neurophysiol.* **61**, 331 (1989).
 - [4] I. Fujita, M. I. K. Tanaka, and K. Cheng, *Nature* **360**, 343 (1992).
 - [5] J. K. Chapin, K. A. Moxon, R. S. Markowitz, and M. A. Nicolelis, *Nat. Neurosci.* **2**, 664 (1999).
 - [6] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events* (John Wiley & Sons, New York, 1966).
 - [7] P. Dayan and L. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, Mass., 2001).
 - [8] S. Shinomoto, Y. Sakai, and S. Funahashi, *Neural Comput.* **11**, 935 (1999).
 - [9] S. Shinomoto, Y. Sakai, and H. Ohno, *Biosystems* **67**, 259 (2002).
 - [10] S. Shinomoto, K. Shima, and J. Tanji, *Neural Comput.* **15**, 2823 (2003).
 - [11] G. R. Holt, W. R. Softky, C. Koch, and R. J. Douglas, *J Neurophysiol.* **75**, 1806 (1996).
 - [12] S. Shinomoto, Y. Miyazaki, H. Tamura, and I. Fujita, *J Neurophysiol.* (in press).

- [13] P. Lansky, R. Rodriguez, and L. Sacerdote, *Neural Comput.* **16**, 477 (2004).
- [14] S. N. Baker and R. N. Lemon, *J Neurophysiol.* **84**, 1770 (2000).
- [15] S. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society, Providence, RI, 2001).
- [16] S. Amari, *IEEE Trans. Inf. Theory* **47**, 1701 (2001).
- [17] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank, *Neural Comput.* **14**, 325 (2002).
- [18] S. Shinomoto, K. Miura, and S. Koyama, *Biosystems* **79**, 67 (2005).
- [19] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer Verlag, Berlin, 1998).
- [20] M. C. Wiener and B. J. Richmond, *J. Neurosci.* **23**, 2394 (2003).
- [21] M. Berry and M. Meister, *J. Neurosci.* **18**, 2200 (1998).
- [22] D. S. Reich, J. D. Victor, and B. W. Knight, *J. Neurosci.* **18**, 10090 (1998).
- [23] J. Keat, P. Reinagel, R. C. Reid, and M. Meister, *Neuron* **30**, 803 (2001).
- [24] J. W. Pillow and E. P. Simoncelli, *Neurocomputing* **52-54**, 109 (2003).
- [25] J. Feng and G. Li, *Neural Comput.* **14**, 621 (2002).
- [26] S. Shinomoto and Y. Tsubo, *Phys. Rev. E* **64**, 041910 (2001).
- [27] M. Tatsuno and M. Okada, *Neural Comput.* **16**, 737 (2004).
- [28] M. Tatsuno and M. Okada, *Neurocomputing* **52-54**, 649 (2003).
- [29] H. Nakahara and S. Amari, *Neural Comput.* **14**, 2269 (2002).